Computing and Informatics, Vol. 27, 2008, 73–92

A FRAMEWORK FOR SEMANTIC INTEROPERABILITY FOR DISTRIBUTED GEOSPATIAL REPOSITORIES

Manoj Paul, Soumya Kanti Ghosh

School of Information Technology
Indian Institute of Technology, Kharagpur
WB 721302, India
e-mail: {mpaul, skg}@sit.iitkgp.ernet.in

Revised manuscript received 11 January 2007

Abstract. Interoperable access of geospatial information across disparate geospatial applications has become essential. Geospatial data are highly heterogeneous – the heterogeneity arises both at the syntactic and semantic levels. Finding and accessing appropriate data in such a distributed environment is an important research issue. The paper proposes a methodology for interoperable access of geospatial information based on Open Geospatial Consortium (OGC) specified standards. An architecture for integrating diverse geospatial data repositories has been proposed using service-based methodology. The semantic issues for discovery and retrieval of geospatial data over distributed geospatial services have also been proposed in the paper. The proposed architecture utilizes the ontological concepts for service description and subsequent discovery of services. An approach for semantic similarity assessment of geospatial services has been discussed.

Keywords: Geospatial ontology, spatial semantics, spatial services, service similarity assessment

1 INTRODUCTION

Sharing geospatial information across diverse geospatial applications has become an important issue. With the advancements of web technology, an increasing number of geospatial data providers want to share data over the web. The utmost requirement to achieve this goal is to build up a methodology for homogeneous access of geospatial information. The diverse nature of geospatial data and lack of standards poses a major challenge for integrating the distributed geospatial data repositories. Since most of the GISs are not originally designed to work in co-operation, several interoperability problems arise while integrating these diverse spatial data sources. Each provider provides its own proprietary data format as well as its specific query language [1]. In the domain of GIS, each dominant GIS platform (e.g., ArcInfo[1], MapInfo[2]) and several database manufacturers (e.g., Oracle[3]) provide data, services and product series based on their own spatial data models, so it is very hard to make for inter-platform data transmission [21]. In addition, geographic resources are designed for a variety of different purposes. Orthogonal directions in the design of geographic resources may affect the semantics of the data they contain and impair their integration [2, 4]. These discrepancies make the integration of different geographic resources significantly complex.

1.1 Service-based Computing

Web Service [31] is a standard-based software technology, which provides programmers and integrators with the possibility to have all existing and developing systems bind together in a refreshing mode. *Web Service* brings interoperation between softwares, which are provided by different providers, running on different operating systems and written in different programming languages. The key features of *Web Service* lie in that it emphasizes interoperation, and implements inter-platform as well as inter-language fast deployment.

Service-oriented architecture (SOA) describes a software architecture that defines the use of loosely coupled software services to support the requirements of business processes and software users [32]. SOA is a paradigm for organizing and utilizing distributed capabilities that may be under the control of different ownership domains. It provides a uniform means to offer, discover, interact with and use capabilities to produce the desired effects. According to the technical standards and implementation specification of *Web Service*, a *Geographical Information Integration* and *Interoperation Platform* design and implemention have been proposed. The design aim of the platform is to provide service integration which have layers as the basic geographical data organization unit, accomplish data retrieval from multiple GIS data sources based on different platforms, proceed on-line overlapping, analysis, visualization of map data.

1.2 Service-based GIS

Major contributions to achieve GIS interoperability comes from *Open Geospatial Consortium* (OGC) [5]. This is an association of government agencies, research

[1] www.esri.com
[2] www.mapinfo.com
[3] www.oracle.com

organizations, software developers, and systems integrators, defining the requirements, standards, and specifications that will support GIS interoperability. The specifications provided by OGC enable syntactic interoperability and cataloguing of geographic information. However, although the OGC defined catalog support discovery, organization and access of geographic information, it does not provide much support for overcoming semantic heterogeneity problem. The semantic heterogeneity is bound to arise in distributed integration platform where service providers and consumers are unaware of each other. The semantic heterogeneity will result in poor discovery of services when exact matching for services is sought. In [22], semantic heterogeneity is defined as the consequence of different conceptualizations and database representations of a real world fact. Keyword-based search can have low recall [2, 6] if different terminology is used and/or low precision if terms are homonymous, or because of their limited possibilities to express complex queries [23]. As a remedy of the problem, the service description should capture enough semantics and an underlying ontological description of the domain can discover similar (rather than exact) services. The semantic of service requests needs to be resolved properly at the Geospatial Service Broker on the context of the service [15, 16].

In this paper, a methodology has been proposed for homogeneous access of geospatial information in a distributed environment. The approach adheres to OGC open standards for service-based geospatial computing. Two standard Web Service techniques proposed by OGC - Web Feature Service (WFS) [7] and Web Map Service (WMS) [9], have been incorporated effectively in developing a service-oriented integrated system for spatial data repositories. Often, effective service discovery requires an extensive search for appropriate services across multiple application domains. Catalogues support discovery, organization, and access of geographic information and thus help the user to find information that exists. The *Catalogue Service* adopts semantic description of service interfaces for similarity assessment of services for discovery purpose. A methodology for the similarity estimation has also been proposed.

The rest of the paper is organized as follows: Section 2 gives some motivating works for spatial data integration both syntactically and semantically. Section 3 gives a brief description of services based geospatial computing and the semantic issues arising therein. Ontology-based discovery and retrieval architecture for integrating distributed geospatial data repositories is discussed in Section 4. Finally, the conclusion is drawn in Section 5.

2 RELATED WORK

Current efforts to integrate geographic information embrace the idea of metadata standards as the key to information sharing and analysis. These include the *Federal Geographic Data Committee* (FGDC) and the *National Spatial Data Infrastructure*

(NSDI)[4], *Geospatial One-Stop*[5], and the *U. S. Geological Survey's The National Map*[6] as well as standards from the *International Standards Organization* (ISO)[7] for geospatial metadata [10]. The NSDI attempts to bring together geographical information sources from all levels of government and other organizations into a single point of entry for easier access to data.

2.1 Integration at Syntactic Level

Several ways of integrating the spatial data repositories have been proposed in the literature. Most current approach for an integrated co-operative interoperable *Spatial Information System* or *Enterprise-GIS* is warehousing the data from heterogeneous repositories in a clean and consistent form, which takes care of both the semantic and syntactic heterogeneities among the data sources. Roth et al. [11] propose a spatial data warehouse based technique and employ middleware technology for data exchange from the spatial data warehouse. However, datawarehouse based approach has several disadvantages keeping in mind the huge volume of data required to be updated regularly. A WebGIS system has been proposed in [34]. It presented a hierarchical component-based WebGIS model referred to as Geo-Union.

There are some works [12] in the geospatial domain using open standards proposed by OGC. An approach is proposed in [12], which uses WFS-based mediation approach with the help of derived wrappers. It provides a multitier client-server architecture and uses standard WFS wrappers to access data. The wrappers are further extended by derived wrappers that capture additional query capabilities. Although it offers an efficient integration and querying mechanism among heterogeneous data repositories, it doesn't offer the effectiveness of service-based technology in heterogeneous system. The architecture makes the data sources tightly coupled with the mediator.

2.2 Integration at Semantic Level

The application of semantic knowledge model for resolving semantic issues in distributed systems has been tried in recent times [33]. The semantic heterogeneity of service description, which poses a challenge for discovering services, has been discussed in [24, 25]. In [24], the use of ontologies for matching service descriptions based on the meanings of the query parameters rather than exact matching is proposed, as well sa a mechanism for sorting the matching services on the basis of degree of matching. *WordNet* [26] has been used for finding the synonyms and hyponyms of the requested parameters. *WordNet* is a lexical database, inspired by current psycholinguistic theories of human lexical memory. English nouns, verbs,

[4] `www.fgdc.gov/nsdi/nsdi.html`
[5] `www.geodata.gov`
[6] `nationalmap.gov`
[7] `www.iso.org`

adjectives and adverbs are organized into synonym sets, each representing one underlying lexical concept. These approaches contribute little for geospatial service discovery, where the spatial relationship is predominant.

Signature matching [27] is an efficient way for component retrieval from software libraries. The method is efficient in the sense that function signatures are automatically generated from the function code. Furthermore, signature matching efficiently prunes down the functions and/or modules that do not match the query, so that more expensive and precise techniques can be used on the smaller set of remaining candidate components. This method can be used for service similarity matching. However, signature matching considers only function types and ignores their behaviors; and two functions with the same signature can have completely opposite behaviors.

In this paper a semantically enhanced SOA architecture using OGC Services has been proposed for integrating divergent geospatial data repositories located geographically at different places. The advantages and benefits of a service method have been fully exploited for designing the architecture. The discovery and retrieval methodology of geospatial services is achieved based on the similarity measurement.

3 SERVICE AND SEMANTICS

The service-based access of geospatial data consider the physical world in terms of features, e.g. a city, a river all are geographic features. Data is accessed over the web with unique feature identifiers with the incorporation of a number of inputs. An example feature request can be *GetFeature(City, CityName, CityState).* Service-based integration approach for geospatial data repositories should consider the semantics of the service descriptions for discovery purpose. The use of diverse terminology for annotating features leads to semantic problem in service discovery. A city can have old name and new name. Lack of proper knowledge base of this information may lead to a poor response to service request. Further, a service request *GetFeature (Airport, CityName)* may give *NULL* result if the city does not have any airport. But, there may be an airport in a neighboring place, which the requester may be interested in. Since there is no way of getting this information by simple searching method, there will be poor response to service request.

3.1 Feature Ontology

The need of a conceptual model for the domain is necessary for capturing the spatial as well as terminological relationship. While terminological similarity can be captured with the use of standard thesauri like *WordNet*[26], spatial relationship requires to be captured with proper understanding of the domain for automated service discovery. Ontology has found its use in information system domain for conceptual modelling purpose. Ontology can be used to describe the semantics of the information sources and to make the content explicit with respect to the integration tasks. It can be defined in many ways that suits the need of its purpose[6].

An ontology **O** is a tuple consisting of the following:

$$O{:=}(C, H_C, R_C, H_R).$$

The concepts C of a domain are arranged in a subsumption hierarchy H_C. Relations R_C exist between concepts. Relations can also be arranged in a hierarchy H_R. An example ontology that captures the university feature is shown in Figure 1.

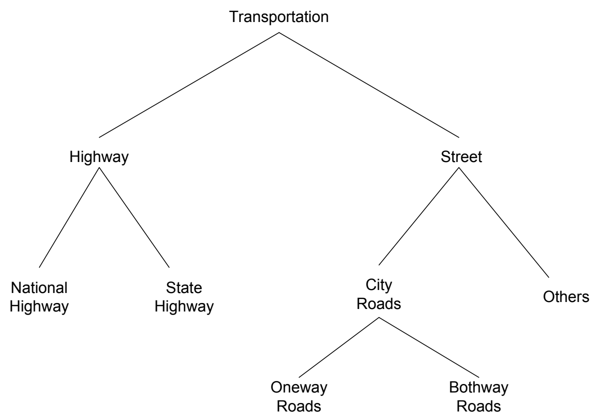

Fig. 1. The hierarchical structure of ontology

There are several languages for describing the ontology in machine-understandable format. We choose OWL [28] for describing the ontology of geospatial features. This language has a well-defined semantics and enables the markup and manipulation of complex taxonomic and logical relations between entities on the Web. It can be used to define the notions of a *Service Profile*, i.e. what the service does, a *Service Model*, i.e. how the service works, and a *Service Grounding*, i.e. how to use the service. Some of the topological relationships are required to be captured in geospatial ontology. They are listed as follows (refer to Figure 2):

- DISJOINT – The boundaries of two spatial features do not intersect.
- TOUCH – The boundaries of two spatial features intersect.
- OVERLAP - There is some common portion in the boundaries of two geospatial features.
- EQUAL – The two features have the same boundary and interior.
- CONTAINS – The boundary of one features is contained within that of the other feature.
- INSIDE – The opposite of CONTAINS. A INSIDE B implies B CONTAINS A.
- COVERED BY – The opposite of COVERS. A COVERED BY B implies B COVERS A.

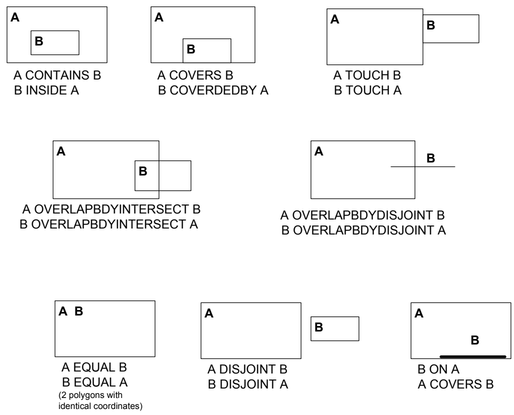

Fig. 2. Spatial relationships as captured in geospatial ontology

Hybrid ontology description [6] approaches have been used to construct ontology of the system (refer to Figure 3). The semantics of each source is described by its own ontology (application ontology). But, in order to make the local ontologies comparable to each other, they are built from a global shared vocabulary (domain ontology). Shared Vocabulary is based on the assumption that members of the domain share a common understanding of certain concepts. Once a shared vocabulary exists, those terms can be used to make the application ontology for a source. The global ontology contains basic terms of a domain, which are combined in the local ontologies in order to describe more complex semantics.

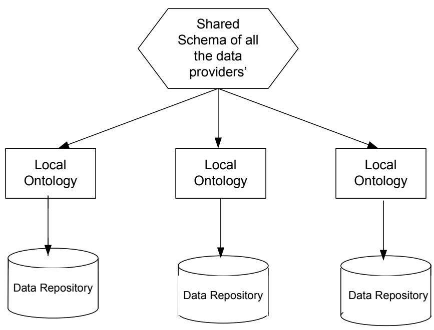

Fig. 3. Structure of hybrid ontology

3.2 Ontology and Location Conceptualization

The ontology description of spatial concepts and their inter-relationship is used to realize the semantics of service request for service discovery in service-based geospatial environments [29, 30]. The matchmaking process, which underlies the ontology-based discovery, is a reasoning process with the goal of deciding, which of the available services are geospatially similar to the request. The main task of the matchmaking process is to resolve semantic heterogeneities between the request and the advertisements. This reasoning perspective emphasizes the need for approaches that go beyond the mere construction of ontologies and involve their use in discovery, evaluating, and combining geospatial information. Semantic matchmaking mechanisms will

- lead to enhanced usability of heterogeneous and distributed geospatial information sources, and
- facilitate the task of automatic service composition.

In general, the geospatial features are a geographical phenomenon. A user searches thematic feature(s)/phenomenon, which has some reference to locations on earth by designating a location associated with that feature. General keyword-based searching methodology for the discovery of a feature location is not adequate for many purposes arising from semantic conflicts. Since this only considers lexicographical matching for searching, it is bound to be inefficient. It yields low precision if homonyms are used for describing geospatial services providing different information. There is a need for methodologies, which can perform imprecise matching (rather than exact matching) of place-name terminology as well [30]. Thus, if the user specifies a place name in a query, then the discovery method should find references to the same or similar places depending on the following geographical closeness properties:

- places that may be referred to by different names, or
- places that may be at different levels of the administrative or topographical hierarchy, or
- places that may be nearby due to connectivity or to some other measure of proximity.

The problems for semantic heterogeneity described in the previous section results from the fact that the terms from a certain vocabulary (e.g. that used in a catalogue) are just words with an implicit meaning (for humans), but without an explicit meaning (which machines are able to understand). Therefore, the usage of concepts that are clearly defined by ontologies to circumvent the ambiguities has to be utilized. If a service request is submitted to the service broker, the broker has to decide which of the registered sources should be provided to answer the request.

The need for service discovery mechanism that recognizes domain-specific terminology has led to various efforts to construct and exploit ontologies that model the associated concepts [14]. In the field of information science, research into thesauri has led to the development of a range of semantic net and thesaurus-mediated information retrieval techniques, for which a variety of semantic closeness metrics have been designed. As indicated above, we are concerned with a conceptualization of place that supports the measurement of locational similarity between named places. The objective is to implement procedures that match a given named location to other named locations that are equivalent or similar in geographical space. It is assumed that a location may refer to any geographical phenomenon, provided that it has been given a name or literal description. The phenomenon, which may have an ontology of its own, may be associated with location ontology.

The similarity measurement method can be performed both on the location and the theme associated with it. Figure 4 illustrates the ontological conceptualization of location and shows how it inherits various attributes and relationship types from geographical concept. A geographical concept has a *Standard Name* (or Preferred Term) and *Alternative Names* (Non-Preferred Terms). These names are associated with alternative spellings, a date of origin and a language. Geographical concepts may be associated with a location defined by a geometry object. The concept *place* is a sub-concept of *geospatial object* concept. In the *place* class, location is specialized by a centroid, defined by latitude and longitude co-ordinates, and spatial relationships are specialized into the *meet*, *overlap* and *partOf* relationships.

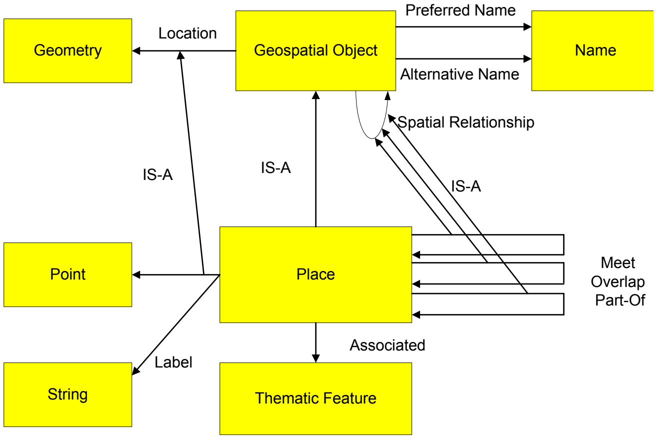

Fig. 4. Locational Ontology capturing the spatial relationships among named places

An instance of the conceptual model of the locational ontology for describing the geographical relatedness of a city with its neighboring places (meets relation) along with the hierarchical containment of it by state/country (*partOf* relation) is shown in Figure 5. The instance model shows the important attributes for geospatial places along with geospatial connectivity with other places. It is worth mentioning here that a place in geospatial space therefore includes physical features of the Earth's surface such as forests, lakes, rivers and mountains, in the natural realm, and cities, counties, roads, and buildings in the human-made environment.

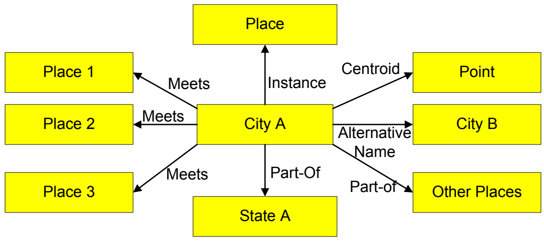

Fig. 5. An instance of locational ontology capturing the geographical closeness of named places

3.3 Similarity Assessment of Geographical Location

The objective of similarity assessment method for geographical locations is to find locations geographically closer to the requested location. Further, we need to rank the results of imprecise geographical queries using location names, which might be global in scope across a wide range of scales. This raises questions of the appropriate types of relationships and semantic attributes to maintain for such applications and leads to the idea of spatial models that record the least amount of information necessary to process the queries. The development of location ontology and generating instances of it by taking care of all the geospatial relationships with other places is the main component for this purpose. Equally important for us is the development of similarity measures that are based on the model of location and that can be used for ranking search results. Potential criteria for assessing locational similarity between a specified place and a candidate place, when searching for information, include the following [29]:

- distance in map or geographical coordinate space between query and candidate;
- number of intervening places;
- spatial inclusion of the candidate within the query place;
- containment of the query place by the candidate;
- containment of candidate within, or overlap of candidate with, regions that contain or overlap the query place;
- boundary connectivity between query and candidate.

The similarity assessment method combines two metrics for geographical closeness, namely, Euclidean distance among query and candidate places and hierarchical distance between query and candidate information sources. While Euclidean distance considers the geographical distance among two places in terms of its centroid, it doesn't consider other aspects of locational closeness such as political, social, etc. Hierarchical distance, on the other hand, can be computed if a conceptual hierarchy is available. The terms within the hierarchy can then be compared by measuring the distance between them along the branches of the corresponding graph. This distance is equal to the number of connecting links in the shortest path in the graph.

The similarity measurement approach among locational entities considers all the non-common parents (at whatever level) of the respective nodes, each of which may have a weight inversely proportional to the depth in the hierarchy. This introduces a measure that the distance between two locations increases with the increasing number of non-common super-concepts. Also the semantic distance between a pair of terms increases in proportion to the number we need to traverse in the ontology to reach that location. Here the similarity measure considers the geographic closeness of places in space irrespective of their structural or terminological similarity. It is assumed that a place is characterized by the sum of the geographical regions, or other parent places, to which it belongs either directly or by inheritance within a hierarchy. A city, for example, may be inside or overlap a county that itself is part of the formal hierarchical administrative subdivision of a nation, which is itself part of a global hierarchy. The hierarchical distance measure is the sum of non-common super-concepts of location of request to those of candidates (refer to Equation (1)):

$$HD_R = \Sigma_{x \in (R.SuperConcepts - C.SuperConcepts)} L_x \tag{1}$$

where L_x represents the hierarchical levels of the individual places within their respective hierarchies. The set of place concepts x are those distinctive super-parts of the query term that belong to it but not to the candidate. Similar approach for calculating the hierarchical distance is used for each candidate and can be derived as Equation (2).

$$HD_C = \Sigma_{y \in (C.SuperConcepts - R.SuperConcepts)} L_y \tag{2}$$

Thus the hierarchical distance of each candidate with respect to the request is given in Equation (3):

$$HD_{R,C} = \alpha * HD_C + \beta * HD_R. \tag{3}$$

The weights α and β provide control over the application of the measure. In general, when applying the hierarchical distance measure, distance between query and candidate increases according to the number of non-common parents, i.e. the distinguishing regions. The level values increase with increasing depth in the hierarchies. This essentially means that there are smaller differences between pairs of places deeper down the hierarchy than there would be higher up. The Euclidean distance between two places is measured with respect to their centroid. Because we are concerned here with applications that may be global in extent, we base the measurement of Euclidean distance on latitude and longitude values for centroid. The two locational distance measures can be combined in a weighted combination referred to as the Spatial Distance (SD) as follows:

$$SD_{R,C} = w_1 * ED_{R,C} + w_2 * HD_{R,C} \tag{4}$$

where w_1 and w_2 are weights of the ED and HD, respectively.

The similarity assessment method can be applied to the searching of geospatial services because the basic unit for geospatial information organization is features and a user searches for features of some geographic location.

3.4 Service Similarity Assessment

For enhancing the semantic discovery of services, we adhere to the standard service description methodology where a service is defined with the following components:

- A signature of the service S_{sig} consisting of the operation names S_{ops}, input/output parameters S_{inp}, S_{out}, respectively. Thus, S_{sig} can be defined as follows (refer Equation (5)):

$$S_{sig} = \{S_{ops} \cup S_{inp} \cup S_{out}\}. \tag{5}$$

  Signature of service gives the static nature of the service.

- Specification of the service, S_{spec}, gives the dynamic aspects of a service. Specifications more precisely characterize the semantics of a component, rather than just its signature.

Given two services S and S' such that $S = \{S_{sig}, S_{spec}\}$ and $S' = \{S'_{sig}, S'_{spec}\}$, the matching between the two service descriptions $Match(S, S')$ can be defined as

$$Match(S, S') = Match(S_{sig}, S'_{sig}) \bigwedge Match(S_{spec}, S'_{spec}). \tag{6}$$

As with *function signature representation* for *Signature Matching*, we can represent service in the same manner by annotating all its inputs, outputs and operation.

Matchmaking Algorithm: In this section we discuss the algorithm for *MatchMaking*. First, we need to generate the candidate service set on the basis of the analysis of requested inputs and advertised inputs as well as requested outputs to those of the advertisements [24]. For this purpose, service request is compared against all the service advertisements from the registry of services. The algorithm is shown in Figure 6.

```
Algorithm:  Matchmaking

Input:  A = Set of all service advertisements
         Q = Service request
         S_cand = φ

Output:  S_cand = Set of all candidate advertisement
         services that can satisfy request
```

```
Begin: Matchmaking
  AddService = TRUE
  For each service advertisement a∈ A do
    For each output of request q_out ∈ Q_out do
      If not Match (a_out, q_out) then
        AddService = FALSE
      EndIf
    EndFor
    If AddService = TRUE
      S_cand = S_cand ∪ a
      Group S_cand in the order of matching degree
    EndIf
  EndFor
End: Matchmaking
```

Fig. 6. MatchMaking algorithm

Scoring Algorithm: Scores are given to each matching services on a discrete scale where the *Exact* match, of course, gets highest score. *PlugIn* is the next candidate in the scoring scale. *Subsumption* comes next, as it can fulfill user requirement partially. *Fail* is given a score of zero, as it cannot satisfy user request at all, hence it can be reported as failure.

$$Exact > PlugIn > Subsumption > Fail$$

The scoring of services in each category is done with the help of ontological scoring process discussed in Section 3.3. This makes use of the similarity score obtained by using semantic scoring process like *Synonym matching*, *Tokenization*, *Abbreviation Expansion*. This is designated as M_i^{sem}. The other scoring method considers ontological similarity and is reffered to as M_i^{ont}.

The candidate services, S_{cand}, are grouped on the discrete scale of four types. They are again ordered on the basis of the similarity score. The services with very low similarity score are pruned out. The algorithm for scoring the services S_{cand} in each groups is shown in Figure 7.

```
Algorithm: Scoring

Input:  S_cand = Set of matched candidate services
        M_threshold = Threshold value for match score

Output:  S' = Set of matched services sorted on score

Begin: Scoring
```

```
For each candidate s ∈ S_cand do
   Calculate domain ontology matching score M_i^sem
   Calculate application ontology matching score M_i^ont
   Set M_i = max {M_i^sem, M_i^ont}

   If M_i < M_threshold do
      S' = S'_cand - s /*services being pruned which are
                        less than threshold score*/
   EndIf

EndFor

Sort S' on similarity score
End:  Scoring
```

Fig. 7. Scoring algorithm

The ranking of the matched services is necessary for preventing the exploitation of the system from forming the following cases – "A service may advertise in more generic way such that it appears that many of the requests can be served by those services. These sorts of services should have lower degree of matching as they do not serve any specific purpose". In the same manner, a requester can make a general request rather than exactly specifying what he expects.

This sort of efficacy can be removed by measuring the degree of matching between the service advertisements and service requests.

3.5 Case Study

In this section, a practical example scenario has been presented for demonstrating the usefulness of the proposed method. Let us consider a user is searching for *Monuments in India*. The similarity measurement approach will find places of type *States*, *Cities* etc. Thus, *TajMahal in Agra* will be retrieved with a score of 100 because *TajMahal* is a *Monument* and *Agra* is within *India*. *Tirupati in Andhra Pradesh* will have score less than 100 because *Tirupati* is a *Temple* and is close to *Monument* concept. Both *Temple* and *Monument* are subclasses of *Artifact* and thus *Temple* could be *PlugIn* with *Monuments*. This way the candidate set is generated for subsequent similarity scoring method. The location ontology hierarchy for the case study is shown in Figure 8. Since our experiments consider the *Artifact* theme, we also utilize an *Artifact Ontology* 9 for similarity measurement on thematic aspect as well.

Let us look at the geographical similarity measurement scenario. For this purpose, it is assumed that a user is interested in searching some artifacts of locations in India. A query like *Palaces in Jaipur* will also retrieve *Palaces in Jodhpur* with

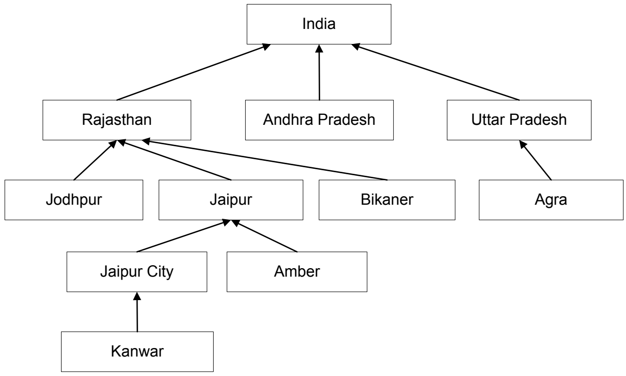

Fig. 8. Ontological hierarchy for *Named Places*

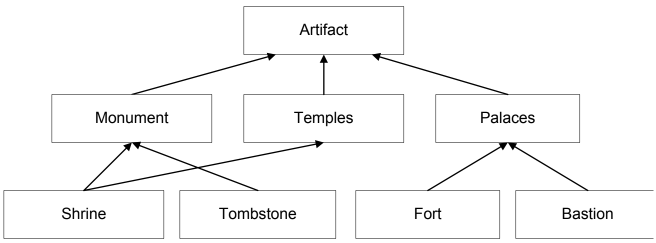

Fig. 9. Artifact ontology

a score close to 100 because *Jodhpur* is near to *Jaipur* and both belong to the state of *Rajasthan*. It should also retrieve *Palaces in Kanwar* with score 100, because Kanwar is a place in Jaipur. Table 1 illustrates the results of ranking retrieved data for a query that requested *Palaces in Jaipur*. In computing the similarity score, those service whose similarity score threshold is greater than 30 have been considered.

Artifact (Theme)	**Location (State)**	**Score**
Palace	Jaipur (Rajasthan)	100
Palace	Kanwar (Rajasthan)	100
Fort	Jaipur (Rajasthan)	98
Palace	Amber (Rajasthan)	94
Monuments	Jaipur (Rajathan)	88
Palace	Jodhpur (Rajasthan)	80
Palace	Agra (Uttar Pradesh)	76
Monuments	Jodhpur (Rajasthan)	72
Temples	Jaipur (Rajasthan)	58
Temples	Bikaner (Rajasthan)	32

Table 1. Ranked result for the query *Palaces in Jaipur*

4 ARCHITECTURE FOR DISTRIBUTED GIS

In this section, an architecture has been proposed for achieving interoperability both at the syntactic and semantic level. Although semantic interoperability is the prime concern, it also takes care of the heterogeneity in data formats. Explication of knowledge by means of ontologies is a possible approach to overcome the problem of semantic heterogeneity. The proposed architecture for service based computing utilizes the ontological descriptions of concepts of the domain of interest for service similarity assessment and hence resolves the semantic heterogeneity problem of the distributed geospatial service platform. Overall architecture of the system is shown in Figure 10. It consists of two main components – *Geospatial Service Providers* (GSP) and *Geospatial Service Broker* (GSB). The GSPs publish the service descriptions in a central *Service Registry* accessible to the GSB. A *Mediator* performs the service registry and brokering job. An *Ontology thesaurus* holds the ontological description of the domain. On receiving a service request from a service consumer, the *Ontology Mapping Engine* matches services in the *Service Registry* with the help of ontological description of the domain.

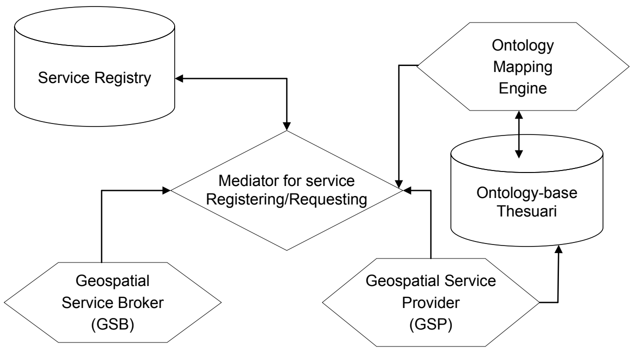

Fig. 10. Architecture for the Service Oriented geospatial computing

In the suggested framework, different data transformation components have been designed for different kinds of data sources (cf. Table 2), namely *shape file*, *GML file*,*Oracle Spatial DBMS*, *PostGIS* etc. The uniformity in heterogeneous GIS data sources is accomplished at the data GSP end. This is accomplished by retrieving data from multiple geospatial data sources based on different platforms, performing on-line overlapping, analysis, visualization of map data, and performing user intercommunication. The *Data Transformation Service* (DTS) is interfaced as wrapper that can communicate data request at multiple data repositories (Figure 11). It accomplishes the function of data retrieval from all sorts of GIS platforms and data format transformation into uniform GML (Geography Markup Language). DTS mainly provides the following GML data transformation abilities of GIS data: *ArcInfo's SHP files*, spatial data stored in *Oracle spatial* and *MapInfo's Tab files*.

Through *Web Service* encapsulation to existing or under developing data transformation modules, DTS provides common transformation interfaces. The processes of analysis of requests, division of requests, delivering small queries are actually designed by using WFS interfaces such as *GetCapabilities*, *DescribeFeatureType*, and *GetFeature*. In [17, 18, 19] more detail on the syntactic interoperability issue for geospatial integration is provided.

Map Data Formate	**Transformers**
Oracle Spatial Database Storage	ORACLEtoGML
ArcInfo File Storage	SHPtoGML
MapInfo File Storage	MAPINFOtoGML

Table 2. Transformers of various geospatial data formats

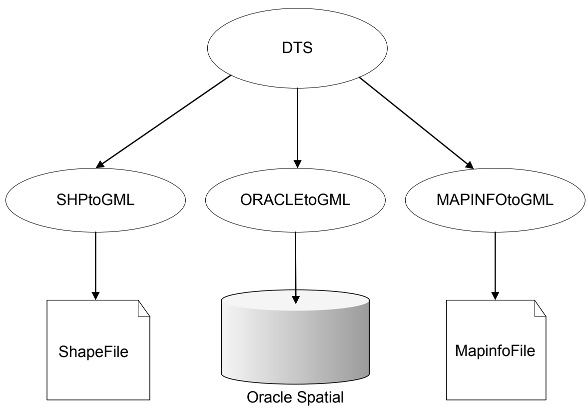

Fig. 11. Heterogeneous data integration platform

5 CONCLUSION

There is an increasing need of geospatial data among a large number of users to share and access the rich geospatial databases that are currently being maintained in several organizations due to wide use of geospatial data and the networked availability of geospatial databases. However, GIS data is immensely heterogeneous both syntactically and semantically. The data is being available in various formats and stored in diverse media (flat files, relational databases). In this paper, service-based methodology has been discussed for integrating distributed geospatial data repositories in adherence to OGC specified open standards.

The paper also describes the central role of a geographic ontology in the development of an integrated information system, which are interoperable semantically.

The complexity arising due to semantic heterogeneity has been addressed. The advantages and benefits of service-based method along with the semantics for services captured in the form of ontology have been fully exploited in designing the integration architecture. Specific issues for geospatial semantics have been addressed while designing the ontology for the geospatial domain. The semantic consideration for geospatial service discovery adds higher efficiency to the system. The similarity assessment method described in the paper is aimed at getting higher precision/recall in service searching. A case study has been proposed as a support for the proposed methodology of semantic interoperability.

REFERENCES

[1] Bishr, Y.: Overcoming the Semantic and Other Barriers to GIS Interoperability. International Journal of Geographical Information Science, 1998, Vol. 12, pp. 299–314.

[2] Cruz, I. F.—Rajendran, A.—Sunna, W.: Handling Semantic Heterogeneities Using Declarative Agreements. Proceedings of ACM GIS, McLean, Virginia, USA, 2002.

[3] Leclercq, E.—Benslimane, D.—Yetongnon, K.: ISIS, A Semantic Mediation Model and an Agent Based Architecture for GIS Interoperability. Proceedings of International Database Engineering and Applications Symposium (IDEAS), 1999, pp. 87–91.

[4] Egenhofer, M. J.: Toward the Semantic Geospatial Web. Proceedings of ACM GIS, November 8–9, McLean, Virginia, USA, 2002.

[5] OGC: Open Geospatial Consortium. `http://www.opengeospatial.org`, 2001.

[6] Wache, H.—Vĕtgele, T.—Visser, U.—Stuckenschmidt, H.—Schuster, G.—Neumann, H.— Hűbner, S.: Ontology-Based Integration of Information – A Survey of Existing Approaches. Proceedings of IJCAI-01 Workshop: Ontologies and Information Sharing, Seattle, WA, 2001, pp. 108–117.

[7] OGC WFS.: Web Feature Service Implementation Specification, version 1.0.0. `http://www.opengeospatial.org/specs/?page=specs`, 2002.

[8] OGC GML: Geography Markup Language (GML) 3.0 Implementation Specification. OGC Recommendation Paper. 2003.

[9] OGC WMS: Web Map Service, Version 1.3. `http://www.opengeospatial.org/specs/?page=specs`, 2004.

[10] Budak, I. A.—Sheth, A.—Ramakrishnan, C.: Geospatial Ontology Development and Semantic Analytics. Handbook of Geographic Information Science, Blackwell Publishing, 2004.

[11] Roth, M. T.— Schwarz, P.: Don't Scrap It, Wrap It! A Wrapper Architecture for Legacy Data Sources. Proceedings of the International Conference on Very Large DataBases (VLDB), 1997.

[12] Boucelma, O.—Garinet, J.—Lacroix, Z.: The VirGIS WFS-Based Spatial Mediation System. Proceedings of ACM CIKM'03, New Orleans, USA, 2003.

[13] Comert, C.: Web Services and National Spatial Data Infrastructure (NSDI). Proceedings of Geo-Imagery Bridging Continents, XXth ISPRS Congress, Istanbul, Turkey Commission 4, 2004.

[14] Klien, E.—Einspanier, U.—Lutz, M.—Hübner, S.: An Architecture for Ontology-Based Discovery and Retrieval of Geographic Information. Proceedings of AGILE Int. Conf. on Geographic Information Science, 2004.

[15] Paul, M.—Ghosh, S. K.: Semantic Interoperability for Enterprise-wide Sharing of Geospatial Information. International Journal of Interoperability in Business Information Systems (IBIS), University of Oldenburg Publication, Germany, 2006, ISSN 1862-6378, Vol. 2, Vol. 2, pp. 21–58.

[16] Paul, M.— Ghosh, S. K.: Geo-Ontology: A Key Component for Service Brokering in Service Oriented Integration of GIS. Proceedings of ISPRS International Symposium on Geospatial Databases for Sustainable Development, Goa, India, 2006.

[17] Paul, M.—Ghosh, S. K.: An Approach for Geospatial Data Management for Efficient Web Retrieval. Proceedings of IEEE International Conference on Computer and Information Technology (CIT), IEEE Computer Society Press, Seoul, Korea, 2006.

[18] Paul, M.—Ghosh, S. K.— Acharya, P. S.: Enterprise Geographic Information System (E-GIS): A Service-based Architecture for Geo-spatial Data Interoperability. Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS), IEEE GRSS Press, Denver, Colorado, USA, July–August, 2006.

[19] Paul, M.—Ghosh, S. K.: An Approach for Service Oriented Discovery and Retrieval of Spatial Data. Proceedings of ACM ICSE International Workshop on Service Oriented Software Engineering (IW-SOSE), ACM Press, Shanghai, China, May, 2006, pp. 88–94.

[20] Paul, M.—Ghosh, S. K.: A Framework for Web-enabled Distributed GIS. Proceedings of International Conference GIS, GPS and Remote Sensing (Map India), New Delhi, India, 2005.

[21] Luo, Y.—Liu, X.—Wang, W.—Wang, X.—Xu, Z.: QoS Analysis on Web Service Based Spatial Integration. Proceedings of International Conference on Grid and Cooperative Computing (GCC), 2004, LNCS 3251, pp. 42–49.

[22] Sahai, A.—Machiraju, V.—Sayal, M.—Jin, L. J.—Casati, F.: Automated SLA Monitoring for Web Services. `http://www.hpl.hp.com/techreports/2002/HPL-2002-191.pdf`, 2002.

[23] Bernstein, A.—Klein, M.: Towards High-Precision Service Retrieval. Proceedings of First International Semantic Web Conference (ISWC), Sardinia, Italy, 2002, pp. 84–101.

[24] Paolucci, M.—Kawamura, T.—Payne, T.—Sycara, K.: Semantic Matching of Web Services Capabilities. Proceedings of 1^{st} International Semantic Web Conference, 2002.

[25] Mahmood, T. S.— Shah, G.— Akkiraju, R.— Ivan, A. A.—Goodwin, R.: Searching Service Repositories by Combining Semantic and Ontological Matching. Proceedings of IEEE International Conference on Web Services (ICWS), 2005.

[26] Miller, G. A.: WordNet: A Lexical Database for the English Language. Communications of ACM, 1983.

[27] Zaremski, A. M.— Wing, J. M.: Signature Matching: A Tool for Using Software Libraries. ACM Transactions on Software Engineering and Methodology, Vol. 4, 1995, No. 2, pp. 146–170.

[28] McGuinness, D. L.—Harmelen, F. V.: OWL Web Ontology Language Overview. W3C Recommendation, 2004.

[29] Jones, C. B.—Alani, H.—Tudhope, D.: Geographical Information Retrieval with Ontologies of Place. Proceedings of the International Conference on Spatial Information Theory: Foundations of Geographic Information Science, 2001, Lecture Notes In Computer Science – 2205, Springer-Verlag, London, pp. 322–335.

[30] Jones, C. B.—Abdelmoty, A. I.—Fu, G.: Maintaining Ontologies for Geographical Information Retrieval on the Web. Proceedings of the OTM Confederated Int. Conf. Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems (ODBASE 03), Springer Verlag, LNCS 2888, pp. 934–951.

[31] Alonso, G.—Casati, F.—Kuno, H.—Machiraju, V.: Web Services: Concepts, Architectures, and Applications. Springer Verlag, 2003.

[32] Channabasavaiah, K.— Holley, K.— Tuggle, E.: Migrating to a Service-Oriented Architecture. IBM DeveloperWorks, December 2003.

[33] Laclavík, M.—Balogh, Z.—Babík, M.— Hluchý, L.: AgentOWL: Semantic Knowledge Model and Agent Architecture. Computing and Informatics, Vol. 25, 2006, No. 5.

[34] Luo, Y. W.— Ding, L.—Wang, X. L.—Wang, W. J.—Xu, Z. Q.: A Hierarchical Component-based WebGIS and Its Key Technologies. Computing and Informatics, Vol. 24, 2005, No. 3.

Manoj Paul is currently pursuing M. Sc. in information technology from Indian Institute of Technology, Kharagpur, India. He is working in the field of spatial data analysis and interoperability. His research interests include spatial data analysis, spatial web services, semantic geospatial web. He is a student member of IEEE.

Soumya Kanti Ghosh is presently working as Assistant Professor in the School of Information Technology, Indian Institute of Technology, Kharagpur, India. Prior to IIT Kharagpur, he worked for Indian Space Research Organization (ISRO), Department of Space, Government of India, for about eight years in the field of satellite remote sensing and GIS. His research interest includes image processing, computer networking, remote sensing and GIS. He is a member of IEEE.